The landscape of soil carbon data: emerging questions, synergies and databases

Progress in
Physical Geography

The landscape of soil carbon data: emerging questions, synergies and databases

Journal:	*Progress in Physical Geography*
Manuscript ID	PPG-19-033.R1
Manuscript Type:	Progress Report
Keywords:	soil carbon data, soil database, wetland carbon, soil chronosequence, soil radiocarbon, long-term ecological research, root traits, soil carbon stabilization, model-data integration, International Soil Carbon Network
Abstract:	Soil carbon has been measured for over a century in applications ranging from understanding biogeochemical processes in natural ecosystems to quantifying the productivity and health of managed systems. Consolidating diverse soil carbon datasets is increasingly important to maximize their value, particularly with growing anthropogenic and climate change pressures. In this progress report, we describe recent advances in soil carbon data led by the International Soil Carbon Network (ISCN) and other networks. We highlight priority research areas requiring soil carbon data, including (i) quantifying boreal, arctic and wetland carbon stocks, (ii) understanding timescales of soil carbon persistence using radiocarbon and chronosequence studies, (iii) synthesizing long-term and experimental data to inform carbon stock vulnerability to global change, (iv) quantifying root influences on soil carbon and (v) identifying gaps in model-data integration. We also describe the landscape of soil datasets currently available, highlighting their strengths, weaknesses and synergies. Now more than ever, integrated soil data are needed to inform climate mitigation, land management and agricultural practices. This report will aid new data users in navigating various soil databases and encourage scientists both to make their measurements publicly available and to join forces to find soil-related solutions.

The landscape of soil carbon data: emerging questions, synergies and databases

Abstract

Soil carbon has been measured for over a century in applications ranging from understanding biogeochemical processes in natural ecosystems to quantifying the productivity and health of managed systems. Consolidating diverse soil carbon datasets is increasingly important to maximize their value, particularly with growing anthropogenic and climate change pressures. In this progress report, we describe recent advances in soil carbon data led by the International Soil Carbon Network (ISCN) and other networks. We highlight priority research areas requiring soil carbon data, including (i) quantifying boreal, arctic and wetland carbon stocks, (ii) understanding timescales of soil carbon persistence using radiocarbon and chronosequence studies, (iii) synthesizing long-term and experimental data to inform carbon stock vulnerability to global change, (iv) quantifying root influences on soil carbon and (v) identifying gaps in model-data integration. We also describe the landscape of soil datasets currently available, highlighting their strengths, weaknesses and synergies. Now more than ever, integrated soil data are needed to inform climate mitigation, land management and agricultural practices. This report will aid new data users in navigating various soil databases and encourage scientists both to make their measurements publicly available and to join forces to find soil-related solutions.

Introduction

Soil carbon is a key component in our understanding of the biosphere's response to global change. There is a long history of soil carbon measurements that, together with other types of soil and ecosystem data, contribute to our understanding of the health and functioning of natural and managed ecosystems (Harden *et al.*, 2018). To better utilize this body of work, the International Soil Carbon Network (ISCN) was formed in 2012 to connect soil carbon researchers and their data. Here, we present recent international efforts consolidating soil carbon data to address urgent soil carbon science questions. We highlight advances in soil databases, led by ISCN or other organizations, to synthesize datasets from diverse sources. Examples include data from boreal, arctic and wetland soils, long-term soil experiments, chronosequences, soil radiocarbon observations and root-soil linkages. These new data will help understand soil carbon stocks, change and vulnerability via syntheses and model-data integration.

What is ISCN?

The International Soil Carbon Network (ISCN) is a science-based network that provides (1) scientific and logistical infrastructure for sharing knowledge, information and data, (2) opportunities for synthesis activities, (3) data products beneficial to stakeholders and scientists and (4) a framework for common scientific protocols and collaborative decision support tools.

Why Soil Carbon?

Soil carbon storage and cycling are measures of a soil's health and ability to cycle nutrients and water as well as to provide services such as food and fiber (Lal, 2004; Banwart *et al.*, 2014). Soil carbon is also directly linked to exchanges of carbon dioxide and trace gases between land-water and land-air systems and therefore is a key component in regulating the global climate system (Ciais *et al.*, 2013). Because soils are a focal point of terrestrial carbon cycling, current research prioritizes quantifying global and ecosystem-specific carbon stocks. In addition to stocks, understanding the processes controlling soil carbon timescales and vulnerability to global change are also critical (Figure 1). These research priorities require diverse data types synthesized across broad scales.

[Insert figure 1]

Why Now?

Land is increasingly under pressure to maintain healthy ecosystems while providing food and fiber to growing human populations. Over one-third of global land surface is currently grazed, forested or cropped (Erb *et al.*, 2007) rendering three quarters or more of the soil carbon down to a meter depth under human management (Harden *et al.*, 2018). Past land management has depleted soil carbon and organic matter (Sanderman, Hengl and Fiske, 2018). However, the re-establishment and buildup of this organic matter through best practices can improve soil productivity and resilience to extreme climate events while also removing carbon dioxide from the atmosphere (Minasny *et al.*, 2017; Batjes, 2019).

Although scientific research on soil carbon has led to numerous sources of data and information, such information is disparate and difficult to access (Harden *et al.*, 2018). Communities interested in making carbon cycle projections or improving agricultural land management need synthesized data to evaluate soil carbon persistence and vulnerabilities to environmental change (Blankinship *et al.*, 2018). With emerging technological advances in data, computing and instrumentation, we see an opportunity to inform and empower land managers with timely, relevant data and information for decision support.

ISCN data holdings

The ISCN database (latest version ISCN3; Nave *et al.*, 2017) contains data from >70,000 soil profiles from a range of data sources, including the United States Department of Agriculture (USDA) and the Northern Circumpolar Soil Carbon Database. More than 200 soil variables are present in the database, including % organic carbon, particle size distribution, pH and % nitrogen. Details of the data types and their calculations can be found at https://iscn.fluxdata.org/data/dataset-information/. Inclusion of a range of supporting measurements (describing geography, soil properties, landform type, etc.) in ISCN makes it possible to investigate soil carbon as part of a dynamic cycle in addition to quantifying stocks. The strengths of the ISCN database include extensive coverage for soil profiles, horizons and depth internationally (with particularly strong representation of the U.S. from USDA data) making ISCN one of the largest, most wide-ranging and diverse repositories of measured soil data.

Recent Advances in ISCN

Shift from template-only to script-based data ingestion

Historically, ISCN has taken a template-based approach to data harmonization where data providers and curators manually input data into the ISCN database. Given that this approach can be labor-intensive and error-prone, ISCN is adding a scripted option for data users and providers. SOC-DRaHR (Soil Organic Carbon Data Rescue and Harmonization Repository; https://github.com/ISCN/SOC-DRaHR; Todd-Brown et al. *in prep)* is a script repository with an associated R package designed to aid in data ingestion and download. SOC-DRaHR also provides a community platform to develop an R library to access and harmonize different data collections.

SOC-DRaHR identifies and downloads soil carbon datasets that are publicly available, provides data harmonization scripts to integrate those data sets into R and provides output scripts for a harmonized data product. In short, these scripts match variable names of the dataset to be ingested with those contained in the ISCN template. SOC-DRaHR is not a data repository or archive but instead an open-source software project that facilitates access to data and harmonizing units and naming conventions across data collections. One limitation of a script-based approach is that it may decrease data user/provider accessibility if they do not have experience with R or other programming languages. To address this, we will keep the template option for users that prefer it.

Led by Katherine Todd-Brown, ISCN hosted two data hackathons (2016 in New Orleans, LA; 2017 in College Station, TX) to train potential contributors and users of ISCN data on our scripted-approach. We also provided guidance and expertise to other science communities building soil or ecological databases (Table 1).

Shift toward open data

The ISCN3 database contains data from sources with varying data-use policies (http://iscn.fluxdata.org/data/dataset-information/data-policy/). In the future, ISCN4 and subsequent versions will only contain data that are open-source under a Creative Commons Attribution (CC-BY) License. The key update under this license will be that requirements of data-provider involvement will be removed but data attribution will be required as before. Previous versions of data bound to sharing restrictions will be retained but only available through ISCN3. ISCN4 will include the open-source data from ISCN3 plus new datasets (Table 1). We consider this open-source shift an important step in making ISCN data easily accessible and usable.

ISCN-led community activities

We held our most-recent all-hands meeting at the American Geophysical Union (AGU) fall meeting in December 2017. The meeting included updates from ISCN as well as breakout groups on root-soil linkages, wetland soil carbon, turnover times of soil carbon and reconciling multi-scale data (http://iscn.fluxdata.org/2018/02/06/summary-of-pre-agu-2017-activities/). We also organized oral and poster sessions at AGU 2017 and 2018. In February 2017, we organized a workshop (Loisel, Malhotra and Phillips, 2017) to discuss and define research and data priorities for soil carbon science and for ISCN. We drafted an article highlighting the converging needs of the soil carbon science and soil health communities and the way forward for ISCN (Harden *et al.*, 2018). ISCN plans to continue to coordinate and host workshops, data hackathons and scientific sessions at international meetings (AGU, European Geophysical Union, etc.).

New datasets and emerging ISCN partnerships

Advances in northern and wetland soil carbon data

Northern peatlands and permafrost soils are rich in carbon that is vulnerable to increased rates of warming and other feedbacks with climate change (Gorham, 1991; Oechel *et al.*, 1993; Frolking, Roulet and Fuglestvedt, 2006; Tarnocai *et al.*, 2009; Schuur *et al.*, 2015). The drivers of soil carbon storage in organic soils can vary considerably relative to mineral soils (Limpens *et al.*, 2008; Loranty *et al.*, 2018; Malhotra *et al.*, 2018; Schuur and Mack, 2018). To better place these soils in a global context, ISCN is including more data from peatlands (Treat *et al.*, 2016, data from C-PEAT; https://github.com/ISCN/soilDataR/blob/master/R/readCPEAT.R) in the next version of the database (ISCN4; Table 1). We will also include Canadian forest soil surveys representing a decade of data (Shaw *et al.*, 2018). Though not always organic soils, these northern forest soils are also expected to undergo warming (Meehl *et al.*, 2007) and provide opportunities for contrasting studies of mineral and organic soils across climate gradients.

Bridging gaps in soil data types

The strengths of the ISCN3 database lie in global survey data that are reported with a range of supporting measurements and are best suited for investigating mechanisms of soil carbon change

(Figure 1). However, other types of data are necessary for carbon stock and vulnerability questions (Figure 1). For example, data from coastal systems, radiocarbon measurements, soil chronosequences, experiments (field manipulations), long-term repeat measurements or root-soil linkages. In an effort to increase our representation of diverse data types, we have built informal (sharing best practices, data harmonization scripts, etc.) or formal (memoranda of understanding) synergies with various groups discussed below.

Coastal wetland carbon: Coastal wetlands are highly productive, and because they form soil as a dynamic response to sea-level rise (Kirwan and Megonigal, 2013) they act as long-term carbon sinks. "Blue Carbon" syntheses have been used to support local greenhouse gas mitigation efforts (Kroeger *et al.*, 2017), to include coastal wetlands in national-scale greenhouse gas inventories (Crooks *et al.*, 2018; Holmquist *et al.*, 2018) and to complete terrestrial-aquatic interface carbon budgets (Najjar *et al.*, 2018). There is a tremendous need for a transparent, well-sourced and living synthesis of coastal carbon stocks. The Coastal Carbon Research Coordination Network (CC-RCN) is currently building such a dataset iteratively: producing standards for data formatting, assisting researchers in creating citable open data releases (Reichman, Jones and Schildhauer, 2011; Wilson *et al.*, 2017) and compiling public data releases into a central data clearing house. CC-RCN personnel are available (until at least 2021) to help providers prepare datasets for submission. To date, the CC-RCN has synthesized data from 3,117 cores from salt marshes, mangroves and tidal freshwater wetlands of the Contiguous United States (Holmquist *et al.*, 2018) and from around the world. ISCN and CC-RCN share lessons-learned on database best practices through workshops and hackathons. In the future, we aspire to formally link our databases through SOC-DRaRH.

Soil radiocarbon data: The International Soil Radiocarbon Database (ISRaD) is an open-source community-based project that brings together soil radiocarbon data and associated datasets (Lawrence *et al.*, 2019). Radiocarbon data are an important tool for understanding the soil carbon cycle and can be used to constrain rates of carbon cycling in models (He *et al.*, 2016) and to assess the timescales and persistence of soil carbon (Sierra *et al.*, 2018). In particular, the application of radiocarbon methodology to improve our understanding of soil carbon dynamics has emphasized the need to conceptualize soils as a consortium of different carbon types, stabilized in soils via a variety of mechanisms. As such there is a growing abundance of soil data collected from specific soil "fractions" that have been physically (e.g., density or particle size separations), chemically (e.g. chemical extractions) or biologically (e.g., soil incubations) partitioned from bulk soil (Poeplau *et al.*, 2018). While these data may provide insight to the nature of a particular soil, it is often challenging to compare fractions across different soils because fractional methods vary widely. ISRaD also seeks to improve our ability to compare soil fractions and standardize fractionation methods, in addition to making soil radiocarbon data more accessible. Data within ISRaD are structured hierarchically and include bulk soil radiocarbon data (approx. 500 sites and 1700 profiles), fractionation schemes (>3600 data points entered), flux measurements (>2100), incubations (>1900), interstitial gases and dissolved organics. Users can add data through a template, which is structured to reflect this hierarchy, or use a scripted approach for larger datasets. In addition to the dataset, ISRaD also offers an associated R package, which includes quality control checks and tools for exploring the data. Ongoing synthesis activities have compiled radiocarbon data from carbon fluxes in the Arctic to look at the potential release of old permafrost carbon (Estop-Aragonés *et al., in review*), from soil incubations to assess rates of fast-cycling soil carbon (Hoyt *et al., in prep*) and from different soil fractions (Heckman *et al., in prep*). Although radiocarbon is the focus of the database, it is not a requirement, allowing the template, data structure and associated tools to be used for other soil-carbon related synthesis efforts. The ISRaD data template builds upon the ISCN template and profile-level soil data will be shared between ISRaD and ISCN.

Soil chronosequence data: Understanding long-term soil carbon dynamics is important for constraining the capacity of soils to store carbon and the spatiotemporal variations in soil carbon related to pedogenic mineralogy. The chronosequence approach has been traditionally used to study the role of time in pedogenesis (Stevens and Walker, 1970). As a result, many chronosequence studies have reported soil carbon data along with other soil and environmental variables. Comparisons of several chronosequences have been used to determine general patterns in soil and ecosystem development and to investigate the effects of other soil-forming factors on carbon, nutrients and mineralogy (Wardle, Walker and Bardgett, 2004). Therefore, a recent effort synthesized data from soil chronosequences with the goal of determining controls of long-term soil carbon dynamics during soil development (Vindušková, Harden, Lawrence, Jackson, *in prep*). The structure of this dataset follows the hierarchical structure of ISCN and draws upon ISRaD in terms of the included variables and tools for data analysis. Upon completion, data from this synthesis will be ingested into the ISCN database.

Experimental and long-term data: Cross-site analysis is a central goal of the Long-term Ecological Research (LTER) program and significant advances have been made in synthesizing cross-site data in hydrology, vegetation dynamics, diversity and climate (Peters *et al.*, 2013). Although soil carbon has been measured at almost all LTER sites as well as at sites from other research networks, cross-network data have, to our knowledge, never been synthesized, compared, modeled or archived in standardized ways across sites (Weintraub *et al.*, 2019). A new synthesis project (Wieder, Lajtha, *et al.*, *in prep*) is addressing this gap by synthesizing long-term soil carbon data not just from LTER sites, but also from National Ecological Observatory Network (NEON), Critical Zone Observatory (CZO), Detritus Input and Removal Treatment (DIRT) and Nutrient Network (NutNet). This project uses a scripted approach similar to ISCN and ISRaD and involves researchers who developed soil models such as MIMICS and CORPSE (Sulman *et al.*, 2014; Wieder *et al.*, 2015), as well as principal investigators who collected the soil carbon data. The model-data synthesis aims to answer questions such as: What roles do microbial and plant community composition play in the transfer of microbial byproducts to persistent soil organic matter (SOM)? How does nitrogen deposition affect SOM composition across a range of climate and mineralogy? Practical implications include outreach to land managers concerned with soil carbon consequences of specific practices.

Linking root traits to soil carbon: Plant root inputs are more likely to be stabilized as long-term soil carbon relative to above ground plant inputs (Jackson *et al.*, 2017; Sokol and Bradford, 2018; Sokol *et al.*, 2018). Despite their recognized importance in soil carbon dynamics, data on root attributes or traits (e.g., root biomass, rooting depth) are severely lacking in soil databases (Harden *et al.*, 2018). Recently, root observations from across the globe have been compiled into the Fine-Root Ecology Database (FRED; Iversen *et al.*, 2017). Whereas FRED version 2.0 includes more than 100,000 root trait observations, it also includes relevant ancillary data such as soil properties, providing an opportunity to harmonize soil and root data. In the past year ISCN held breakout group discussions and a workshop to develop a framework linking root traits with soil carbon across the globe (Malhotra, Sihi and Iversen, 2018). The root trait working group will continue their efforts in 2019, focusing on the three main stages of root-soil interactions, namely rhizosphere engineering by living roots, root inputs to soil organic matter via turnover and the decay of root necromass throughout the soil profile.

Mechanisms of soil carbon storage and stability: In coordination with ISCN, the US Geological Survey (USGS) and the US Department of Agriculture recently supported a series of USGS Powell Center workshops targeted towards improving our understanding of mechanisms controlling soil carbon storage and stability. Several products were derived from these workshops, including an exploration of how soils measurements, models and theories are linked in order to better integrate rapidly expanding soil research efforts (Blankinship *et al.*, 2018) and a reevaluation of soil carbon controls using existing databases (Rasmussen *et al.*, 2018). The results of these workshops highlight the critical importance of including ancillary soil data in soil carbon syntheses and provide further opportunity to better coordinate future soil measurements with models and theory.

Model-data integration: Soil data synthesis efforts strive to inform model development and validation. Model evaluation is an important goal of the International Land Model Benchmarking Project (ILAMB; Collier *et al.*, 2018). ISCN participated in ILAMB's soil organic carbon working group (Oak Ridge National Laboratory, October 2018) to develop ISCN-derived data products that would be useful for model benchmarking.

Beyond benchmarking, there is a growing potential to use synthesized datasets for model-data integration to develop our understanding of soil carbon dynamics (Bloom *et al.*, 2016; Luo *et al.*, 2016). Model-data integration activities can help determine model structures and parameterizations that are consistent with observations of carbon stocks, soil ages (radiocarbon data), above- and belowground litter inputs and local conditions (soil texture, moisture and temperature), weighted according to measurement error. Advances in computing power and algorithm development allow model calibration and evaluation across very large datasets, facilitating our capacity to simulate soil processes regionally and globally. A key request from the model-data integration community is that soil databases include clear quantification of all sources of measurement error (to allow for Bayesian statistics). Additionally, if point data have been converted to gridded products, the upscaling error is key for model-data integration.

Navigating the landscape of soil data

The landscape of soil data is complicated and contains a range of databases representing different regions and variables (Figure 2). To a new data user (e.g., a graduate student), it may be daunting to select the right dataset to answer a research question or the best database to target for their data contributions. One of ISCN's missions is to inform data users of the strengths and weaknesses of each database and circumvent issues related with multiple soil databases that are difficult to harmonize. Our recent synergies with CC-RCN, ISRaD, chronosequences and the LTER/NEON/CZO data syntheses were therefore initiated with the intention of sharing information on best practices, standardizing controlled vocabularies and providing resources such as R scripts to ingest or harmonize data.

Additionally, ISCN and the International Soil Reference and Information Centre (ISRIC) developed a formal agreement to ensure that ISCN soil profile data are fed into ISRIC's spatially-extensive database on a regular basis, following a screening for possible duplicate profiles (Ribeiro, Batjes and Van Oostrum, 2018). If a user is interested in global carbon stocks, they may use the entire WoSIS (World Soil Information Service) database (Batjes *et al.*, 2017) or its derived products (SoilGrids250m; Hengl *et al.*, 2017). However, if a user is interested in abiotic or mechanistic controls of soil carbon, ISCN may be more appropriate, as it provides more ancillary data on soil properties and ecology than the ISRIC database.

[Insert figure 2]

Future directions

In the short-term, our goal is to provide data infrastructure that enables interoperability not just between ISCN data sources but also across the synthesis efforts mentioned here. This is a non-trivial task, but the community is ready and the need for harmonized soil datasets is clear.

In the longer-term, in addition to maintaining the aforementioned data and infrastructure, we would also like to consolidate new data sources and types. Most urgently, given that managed soil extent exceeds that of unmanaged soils globally (Harden *et al.*, 2018), ISCN would like to include more data from agricultural and other managed systems. We hope to continue our discussions with entities such as FarmOS (https://farmos.org/) and CIRCASA (Coordination of International Research Cooperation on soil CArbon Sequestration in Agriculture; https://www.circasa-project.eu/) to consolidate agricultural data into a central repository. This first step is necessary to link management practices to resulting soil properties.

Activities summarized in this report highlight emerging priorities within soil carbon science. We especially highlight recent advances in high-latitude soils and at the terrestrial-aquatic interface as well as in experimental, long-term, chronosequence or radiocarbon data. In a complex landscape of soil carbon data and applications, ISCN and our partners strive to provide resources, data and opportunities for disparate soil carbon communities to exchange ideas and solutions. Promoting healthy soils and finding creative solutions for climate change mitigation and adaptation will require collaborations among land managers, policy makers and scientists. We hope our report will serve as a call for input, not only of data, but also of best-practices, code, resources and ways forward, from other soil carbon-relevant entities, databases and networks.

Literature cited

Banwart, S. *et al.* (2014) 'Benefits of soil carbon: Report on the outcomes of an international scientific committee on problems of the environment rapid assessment workshop', *Carbon Management*, 5(2), pp. 185–192. doi: 10.1080/17583004.2014.913380.

Batjes, N. H. *et al.* (2017) 'WoSIS: Providing standardised soil profile data for the world', *Earth System Science Data*, 9(1), pp. 1–14. doi: 10.5194/essd-9-1-2017.

Batjes, N. H. (2019) 'Technologically achievable soil organic carbon sequestration in world croplands and grasslands', *Land Degradation and Development*, 30(1), pp. 25–32. doi: 10.1002/ldr.3209.

Blankinship, J. C. *et al.* (2018) 'Improving understanding of soil organic matter dynamics by triangulating theories, measurements, and models', *Biogeochemistry*, 140(1), pp. 1–13. doi: 10.1007/s10533-018-0478-2.

Bloom, A. A. *et al.* (2016) 'The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times', *Proceedings of the National Academy of Sciences*, 113(5), pp. 1285–1290. doi: 10.1073/pnas.1515160113.

Ciais, P. *et al.* (2013) 'Attributing the increase in atmospheric CO2 to emitters and absorbers', *Nature Climate Change*. Nature Publishing Group, 3(10), pp. 926–930. doi: 10.1038/nclimate1942.

Collier, N. *et al.* (2018) 'The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation', *Journal of Advances in Modeling Earth Systems*, pp. 2731–2754. doi: 10.1029/2018MS001354.

Crooks, S. *et al.* (2018) 'Coastal wetland management as a contribution to the US National Greenhouse Gas Inventory', *Nature Climate Change*. Springer US, 8(December). doi: 10.1038/s41558-018-0345-0.

Erb, K. H. *et al.* (2007) 'A comprehensive global 5 min resolution land-use data set for the year 2000 consistent with national census data', *Journal of Land Use Science*, 2(3), pp. 191–224. doi:

10.1080/17474230701622981.
Frolking, S., Roulet, N. and Fuglestvedt, J. (2006) 'How northern peatlands influence the Earth's radiative budget: Sustained methane emission versus sustained carbon sequestration', *Journal of Geophysical Research: Biogeosciences*, 111(1), pp. 1–10. doi: 10.1029/2005JG000091.
Gorham, E. (1991) 'Northern peatlands: role in the carbon cycle and probable responses to climatic warming', *Ecological Applications*, 1(2), pp. 182–195. Available at: http://www.esajournals.org/doi/abs/10.2307/1941811 (Accessed: 15 June 2013).
Harden, J. W. *et al.* (2018) 'Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter', *Global Change Biology*, 24(2). doi: 10.1111/gcb.13896.
He, Y. *et al.* (2016) 'Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century', *Science*, 353(6306), pp. 1419–1424.
Hengl, T. *et al.* (2017) 'SoilGrids250m: Global gridded soil information based on machine learning', *PLOS ONE*. Public Library of Science, 12(2), p. e0169748. Available at: https://doi.org/10.1371/journal.pone.0169748.
Holmquist, J. R. *et al.* (2018) 'Accuracy and Precision of Tidal Wetland Soil Carbon Mapping in the Conterminous United States', *Scientific reports*, 8(1), p. 9478. doi: 10.1038/s41598-018-26948-7.
Iversen, C. M. *et al.* (2017) 'Fine-Root Ecology Database (FRED): A Global Collection of Root Trait Data with Coincident Site, Vegetation, Edaphic, and Climatic Data, Version 1'. doi: 10.3334/cdiac/ornlsfa.005.
Jackson *et al.* (2017) 'The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls', *Annual Review of Ecology, Evolution, and Systematics*, 48(1), pp. 419–445. doi: 10.1146/annurev-ecolsys-112414-054234.
Kirwan, M. L. and Megonigal, J. P. (2013) 'Tidal wetland stability in the face of human impacts and sea-level rise', *Nature*, 504(7478), pp. 53–60. doi: 10.1038/nature12856.
Kroeger, K. D. *et al.* (2017) 'Restoring tides to reduce methane emissions in impounded wetlands: A new and potent Blue Carbon climate change intervention', *Scientific Reports*. Springer US, 7(1), pp. 1–12. doi: 10.1038/s41598-017-12138-4.
Lal, R. (2004) 'Soil carbon sequestration impacts on global climate change and food security.', *Science*, 304(5677), pp. 1623–7. doi: 10.1126/science.1097396.
Lawrence, C. R. *et al.* (2019) 'An open source database for the synthesis of soil radiocarbon data: ISRaD version 1.0', *Earth System Science Data Discussions*, pp. 1–37. doi: 10.5194/essd-2019-55.
Limpens, J. *et al.* (2008) 'Peatlands and the carbon cycle: from local processes to global implications – a synthesis', *Biogeosciences*, 5(5), pp. 1475–1491. doi: 10.5194/bg-5-1475-2008.
Loisel, J., Malhotra, A. and Phillips, C. (2017) 'A new platform for managing soil carbon and soil health', *Eos*, 98(11). doi: https://doi.org/10.1029/2017EO080753.
Loranty, M. M. *et al.* (2018) 'Reviews and syntheses: Changing ecosystem influences on soil thermal regimes in northern high-latitude permafrost regions', *Biogeosciences*, 15(17), pp. 5287–5313. doi: 10.5194/bg-15-5287-2018.
Luo, Y. *et al.* (2016) 'Toward more realistic projections of soil carbon dynamics by Earth system models', *Global Biogeochemical Cycles*, 30, pp. 40–56. doi: 10.1002/2015GB005239.Received.
Malhotra, A. *et al.* (2018) 'Post-thaw variability in litter decomposition best explained by microtopography at an ice-rich permafrost peatland', *Arctic, Antarctic, and Alpine Research*. Taylor & Francis, 50(1), pp. 1–10. doi: 10.1080/15230430.2017.1415622.

Malhotra, A., Sihi, D. and Iversen, C. (2018) ‘The Fate of Root Carbon in Soil: Data and Model Gaps’, *Eos*, 99. doi: 10.1029/2018EO112593.
Meehl, G. A. *et al.* (2007) ‘IPCC, 2007: Climate Change 2007: the physical science basis.’, in *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, U.K.: Cambridge University Press; 2007, pp. 747–846. doi: 1452cb7a-9f93-44ea-9ac4-fd9f6fd80a07.
Minasny, B. *et al.* (2017) ‘Soil carbon 4 per mille’, *Geoderma*. The Authors, (292), pp. 59–86. doi: 10.1016/j.geoderma.2017.01.002.
Najjar, R. G. *et al.* (2018) ‘Carbon Budget of Tidal Wetlands, Estuaries, and Shelf Waters of Eastern North America’, *Global Biogeochemical Cycles*, 32(3), pp. 389–416. doi: 10.1002/2017GB005790.
Nave, L. *et al.* (2017) ‘International Soil Carbon Network (ISCN) Database v3-1’. doi: 10.17040/ISCN/1305039.
Oechel, W. C. *et al.* (1993) ‘Recent change of Arctic tundra ecosystems from a net carbon dioxide sink to a source’, *Nature*, 361(6412), pp. 520–523. doi: 10.1038/361520a0.
Peters, D. *et al.* (2013) *Long-Term Trends in Ecological Systems: A Basis for Understanding Responses to Global Change. U.S.D.A. Agricultural Research Service, Technical Bulletin No. 1931*.
Poeplau, C. *et al.* (2018) ‘Isolating organic carbon fractions with varying turnover rates in temperate agricultural soils – A comprehensive method comparison’, *Soil Biology and Biochemistry*. Elsevier, 125(April), pp. 10–26. doi: 10.1016/j.soilbio.2018.06.025.
Rasmussen, C. *et al.* (2018) ‘Beyond clay: towards an improved set of variables for predicting soil organic matter content’, *Biogeochemistry*, 137(3), pp. 297–306. doi: 10.1007/s10533-018-0424-3.
Reichman, O. J., Jones, M. B. and Schildhauer, M. P. (2011) ‘Challenges and opportunities of open data in ecology’, *Science*, 331(6018), pp. 703–705. doi: 10.1126/science.1197962.
Ribeiro, E., Batjes, N. H. and Van Oostrum, A. J. M. (2018) *‘World Soil Information Service (WoSIS) - Towards the standardization and harmonization of world soil data’. Procedures Manual 2018. ISRIC Report 2018/01, ISRIC - World Soil Information, Wageningen*. doi: http://dx.doi.org/10.17027/isric-wdcsoils.20180001.
Sanderman, J., Hengl, T. and Fiske, G. J. (2018) ‘Soil carbon debt of 12,000 years of human land use’, *Proceedings of the National Academy of Sciences*, 115(7), p. 201800925. doi: 10.1073/pnas.1800925115.
Schuur, E. A. G. *et al.* (2015) ‘Climate change and the permafrost carbon feedback’, *Nature*, 520(7546), pp. 171–179. doi: 10.1038/nature14338.
Schuur, E. A. G. and Mack, M. C. (2018) ‘Ecological response to permafrost thaw and consequences for local and global ecosystem services’, *Annual Review of Ecology, Evolution, and Systematics*, 49(1), pp. 279–301. doi: 10.1145/1999030.1999036.
Shaw, C. *et al.* (2018) ‘A Canadian upland forest soil profile and carbon stocks database’, *Ecology*, 99(4), p. 989. doi: 10.1002/ecy.2159.
Sierra, C. A. *et al.* (2018) ‘Soil Organic Matter Persistence as a Stochastic Process: Age and Transit Time Distributions of Carbon in Soils’, *Global Biogeochemical Cycles*, 32(10), pp. 1574–1588. doi: 10.1029/2018GB005950.
Sokol, N. W. *et al.* (2018) ‘Evidence for the primacy of living root inputs, not root or shoot litter, in forming soil organic carbon’, *New Phytologist*, pp. 233–246. doi: 10.1111/nph.15361.
Sokol, N. W. and Bradford, M. A. (2018) ‘Microbial formation of stable soil carbon is more

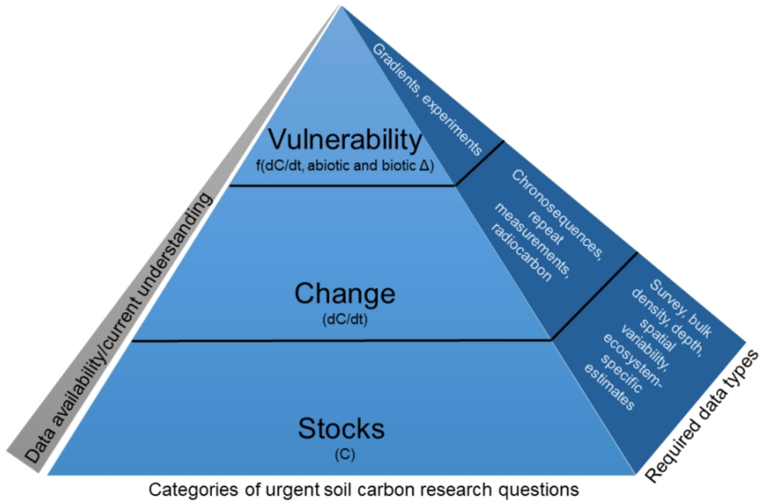
efficient from belowground than aboveground input', *Nature Geoscience*. Springer US, 12(January). doi: 10.1038/s41561-018-0258-6.
Stevens, P. R. and Walker, T. W. (1970) 'The Chronosequence Concept and Soil Formation', *The Quarterly Review of Biology*, 45(4), pp. 333–350. doi: 10.1086/406646.
Sulman, B. N. *et al.* (2014) 'Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO 2', *Nature Climate Change*, 4(12), pp. 1099–1102. doi: 10.1038/nclimate2436.
Tarnocai, C. *et al.* (2009) 'Soil organic carbon pools in the northern circumpolar permafrost region', *Global Biogeochemical Cycles*, 23(2), p. GB2023. doi: 10.1029/2008GB003327.
Treat, C. C. *et al.* (2016) 'Effects of permafrost aggradation on peat properties as determined from a pan-Arctic synthesis of plant macrofossils', *Journal of Geophysical Research: Biogeosciences*, 121(1), pp. 78–94. doi: 10.1002/2015JG003061.
Wardle, D., Walker, L. and Bardgett, R. (2004) 'Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences', *Science*, 305(5683), pp. 509–513. doi: 10.1126/science.1098778.
Weintraub, S. R. *et al.* (2019) 'Leveraging Environmental Research and Observation Networks to Advance Soil Carbon Science Journal of Geophysical Research : Biogeosciences', (2016), pp. 1–9. doi: 10.1029/2018JG004956.
Wieder, W. R. *et al.* (2015) 'Representing life in the Earth system with soil microbial functional traits in the MIMICS model', *Geoscientific Model Development*, 8(6), pp. 1789–1808. doi: 10.5194/gmd-8-1789-2015.
Wilson, G. *et al.* (2017) 'Good enough practices in scientific computing', *PLOS Computational Biology*, 13(6), p. e1005510. doi: https://doi.org/10.1371/ journal.pcbi.1005510.

Table. 1 New data and research synergies within ISCN and across several soil working groups.

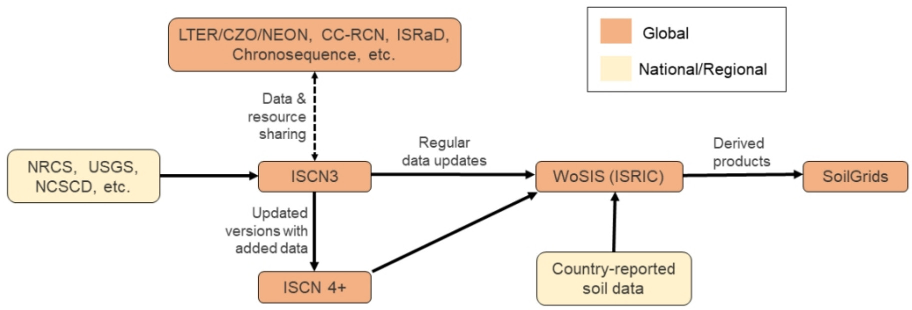
Dataset name	Dominant data type	Number of observations	Attributes	Geographical coverage	URL
ISCN 3	Survey	70k+ profiles	>200	Global but US-heavy	https://iscn.fluxdata.org/data/dataset-information/
Shaw et al. 2018	Canadian forest soils	3000 profiles	60	Canada	Included in ISCN
Treat et al. 2015	Peatland	500 cores	30	Global	Included in ISCN
C-PEAT	Peatland	82 cores	10	Global	Included in ISCN
Coastal Carbon RCN	Wetland carbon	3000 profiles	131	Global but US-heavy	https://github.com/Smithsonian/CCRCN-Data-Library
ISRaD	Radiocarbon	1700 profiles	>250	Global	www.soilradiocarbon.org
LTER SOM	Experimental, repeat measurements	140 locations	170	Global but US-heavy	https://lter.github.io/som-website/
FRED	Belowground trait data	105k root trait observations	300	Global	https://roots.ornl.gov/
ISRIC (WoSIS)	Survey	150k profiles	24	Global	https://www.isric.org/explore/wosis/accessing-wosis-derived-datasets

Figure 1. **Key questions in soil carbon science and corresponding data requirements**. Emergent questions about soil carbon vulnerability (response to abiotic/climate or biotic/land cover change, management, disturbance, etc.) must be underpinned by questions of soil carbon change (timescales, persistence and stability, factors controlling microbial access, quality and fraction) which in turn are rooted in questions of carbon stocks (spatial variability, ecosystem-specific storage, depth variation, etc). Synthesis efforts described in this paper represent a range of data/efforts directed toward addressing each of these knowledge gaps. For example, research questions on soil carbon vulnerability may utilize data from experimental manipulations, plant-trait databases; carbon change questions from databases such as ISCN3, chronosequence and radiocarbon syntheses; and carbon stock questions from global or ecosystem-specific survey data.

Figure 2. **Navigating the landscape of soil data**: the ISCN3 database and its current link to other large soil databases. ISCN3 comprises various independent data sources that are globally extensive but with a strong U.S. focus. Data sources include Natural Resources Conservation Service (NRCS), United States Geological Survey (USGS), The Northern Circumpolar Soil Carbon Database (NCSCD), etc. ISCN publishes a new data version periodically (e.g., ISCN4 will contain new northern and peatland data). In turn, ISCN data are regularly ingested into the World Soil Information Service database (WoSIS), a larger database focused on nationally-reported profile data. Global gridded products such as SoilGrids are derived from profiles held in WoSIS, a set of environmental co-variates and digital soil mapping. Lastly, ISCN maintains synergies with various other data synthesis groups (e.g., ISRaD, CC-RCN, LTER; described in text) that encompass data types not well-represented by ISCN (radiocarbon, coastal carbon, experimental manipulations, etc).



338x190mm (96 x 96 DPI)



338x190mm (96 x 96 DPI)